COVER SHEET

Skitmore, Martin (2002) Predicting the probability of winning sealed bid auctions: a comparison of models. Journal of the Operational Research Society 53(1):pp. 47-56.

THE PROBABILITY OF TENDERING THE LOWEST BID IN SEALED BID AUCTIONS: AN EMPIRICAL ANALYSIS OF CONSTRUCTION CONTRACT DATA

Revised paper for

Journal of the Operations Research Society

Martin Skitmore
Professor of Construction Management and Economics
School of Construction Management and Property
Queensland University of Technology
Gardens Point
Brisbane Q4001
Australia

Version 2 (16 May 2006)

THE PROBABILITY OF TENDERING THE LOWEST BID IN SEALED BID AUCTIONS: AN EMPIRICAL ANALYSIS OF CONSTRUCTION CONTRACT DATA

ABSTRACT

This paper is concerned with predicting the probability of tendering the lowest bid in sealed bid auctions. Four of the leading models from the bidding literature are shown to be subsumed within a general model – differing only in their method of parameter estimation. These models are then tested relative to the equal probability model by an empirical analysis of a large sample of real construction contract bidding data via *all-in*, *one-out* and *one-on* sample frames. A binomial test is used to measure the ability to predict the **identity** of the lowest bidders and the average logscore is used to measure the ability to predict the **probability** of each bidder being the lowest. Optimal cut-off criterion values are determined for defining the minimum size of dataset needed for disaggregating bidders. The work also highlights the importance of (1) the treatment of new entrants and general shortage of data on individual bidders, and (2) the treatment of predicted ties.

Keywords: Bidding models, bidding theory, construction contracts, empirical tests, predicted probability, probability of lowest bid, sealed bid auctions, tendering theory, binomial test, logscore test.

INTRODUCTION

The probability of individual contestants winning a bidding auction can be a useful piece of information for many people, not least the contestants themselves. Potential bidders can utilise this information to decide in which auctions to participate, when to try to obtain an invitation to participate, whether to enter a bid and, if so, the dollar value of the bid. Similarly, the auctioneer can also utilise the information in deciding when and how to hold the auction, how many and which bidders to invite, and the criterion for determining the winner.

Auction bidding is about the price to quote in a sealed auction. Most of the literature on the subject is concerned with setting a price, *x*, so that the probability, *Pr(x)*, of winning the auction reaches some desired level. Several methods have been proposed for predicting *Pr (x)*, and these have been subject to quite lengthy, but yet inconclusive, discussion based on the theoretical merits of each method. To date, there have been no empirical tests applied, presumably due to the lack of development of appropriate tests. Recent work by Wallace and Patrick (1993) and Dowe *et al* (1996), however, has demonstrated the use of a logarithmic scoring function for probabilistic predictions and this, together with a new form of binomial test, has cleared the way for the tests described in this paper. These methods are applied to four sets of construction contract bidding data, with some adjustments due to the limitations of the data.

In concentrating exclusively on empirical testing, the work described in this paper is rather less ambitious than previous studies of auction bidding in restricting the scope of the study to the prediction of *Pr(x)* without specifying any related functions. On the other hand, bearing in mind the wide range of potential uses mentioned above, for both bidders and auctioneers, it is possible that the results may have a greater potential breadth of application than normally found in the bidding literature. This paper is presented in the context of construction contract bidding. Although there may be some idiosyncrasies in the procedural details involved, generalisation to other domains should be a relatively simple matter.

PREVIOUS EMPIRICAL STUDIES

The competitive pressures in the construction industry, it has been said, are probably more intense than in any other industry (Park, 1972:24.1). In the presence of such pressures it is not altogether surprising that, judging from the reported attitude of some companies, competitive bidding is less of a competition based on costs or profit margins, than a lottery in which the inherent uncertainty of the process decides the winner (Whittaker, 1970). Indeed, according to McCaffer (1976) there is "... substantial evidence that existing bidding processes are little more than random". Pim's (1974) analysis of the number of projects awarded to four USA construction companies indicates that the average number of projects acquired is generally proportional to the reciprocal of the average number of bidders competing - the proportion that would be expected to be won by pure 'chance' alone (Table 1). This suggests an extremely simple ‘equal probability’ model in which the expected probability of entering the lowest bid in a k-size auction, that is, an auction in which k bidders enter bids, is the reciprocal of k.

Several formulations have been advanced that claim to offer a theoretical improvement on Pim’s equal probability model, none of which have yet been tested empirically. Of these, four main approaches are considered. These comprise Friedman (1956), Gates (1967), Carr

(1982) and Skitmore (1991). As will be shown, all are based on the same statistical model but differ in their detailed assumptions of its specification.

BIDDING MODELS

Let x_i be a bid entered by bidder i for a sealed bid auction and let $x_i \sim f_i\ (\mu_i, \sigma_i)$ where μ_i and σ_i represent the various location and scale parameters respectively in the probability distribution, f_i, from which the ith bid is drawn. If, for a particular auction, the k bids are each treated as continuous random variables with joint probability density function $f(x_1, x_2, \ldots, x_k)$, then the probability, $Pr(x_1$ lowest), that bidder 1 enters the lowest bid, is

$$\Pr(x_1 \, lowest) = \Pr(x_1 < all\ x_i, i \neq 1) = \int_{-\infty}^{x_1} \int_{x_1}^{\infty} \int_{x_1}^{\infty} \ldots \int_{x_1}^{\infty} f\,(x_1, x_2, \ldots, x_k)\ dx_k\, dx_{k-1} \ldots dx_1 \qquad \textbf{\textit{(1)}}$$

Assuming independence, then it follows that

$$\Pr(x_1\ lowest) = \int_{-\infty}^{x_1} f_1(x_1) \prod_{i=2}^{k} \left[\int_{x_1}^{\infty} f_i(x_i)\ dx_i \right] dx_1 \qquad \textbf{\textit{(2)}}$$

In this case, bidder 1 was chosen as the reference bidder although ***(1)*** and ***(2)*** can clearly be modified for other reference bidders in the auction to obtain $\Pr(x_2 lowest)$, $\Pr(x_3 lowest)$, etc. For convenience, however, we will continue to refer to bidder 1 as the reference bidder in the notation.

ESTIMATION METHODS TO BE TESTED

The five estimation methods investigated comprised: (1) the equal probability method; (2) Friedman's (1956) method; (3) Gates' (1967) method; (4) Carr's (1982) method; and (5) Skitmore's (1991) method. These are summarised in Table 2 and described in detail below including the necessary modifications that had to be made due to the limitations of the data and nature of the analysis. All assume independence of bids and therefore ***(2)*** applies.

(1) Equal probability method

The equal probability method, as implied by Pim, directly predicts each bidder's probability of being the lowest bidder as $1/k$, i.e. $\Pr(x_1)=\Pr(x_2)= ... = 1/k$ for a given k-size auction. The same result is obtained when all bids are identically and independently distributed, i.e., where $f_1(\ ) = f_2(\ ) = \ldots, = f_k(\ )$, $\mu_1 = \mu_2 = \ldots , = \mu_k$, and $\sigma_1 = \sigma_2 = \ldots , = \sigma_k$

This is essentially the control model representing chance. A good estimation method should, of course, outperform the equal probability method by definition. Of course, it is not possible to predict winners this way as all are tied.

(2) Friedman's method

Friedman's approach is to transform x_i by dividing by the reference bidder's cost estimates, c_1, i.e.,

$$x_i^F = x_i/c_1 \qquad \textbf{\textit{(3)}}$$

the shape and other distribution parameters for x_i^F being estimated from the frequency distribution of the x_i^F ratios for each competitor. x_1^F is however assigned the arbitrary parameter values $\mu_1 = x^*/c^*$ and $\sigma_1^2 = 0$, where x^* and c^* are the reference bidder's bid and cost estimate for the next auction.

Friedman's approach relies heavily on the availability of data, the theoretical density functions being be fitted provided there are data for “enough previous contracts” (Friedman, 1956:107). A range of criterion values was used to determine this, ie., q=1,2,...,30, where q denotes the minimum number of previous bidding encounters between the reference bidder and a specific competitor. Where the actual number of previous bidding encounters between the reference bidder and a specific competitor was less than q, the probability was estimated as the mean 'success', $\overline{p}$, of the reference bidder against all other bidders, ie.,

$$\overline{p} = \frac{1}{n'}\sum_{i'=1}^{n'} p(i') \qquad \textbf{\textit{(3a)}}$$

where $p(i')$ is the probability of the reference bidder's cost estimate being less than bidder i's bid, $p(i')$ being estimated by the ratio of the number of previous auctions where the reference bidder's cost estimates were less than bidder i's bids to the total number of auctions where the reference bidder bid against bidder i. Where no previous meetings of a pair of bidders had taken place, each bidder in the pair was assigned a 0.5 probability of underbidding the other.

To enable *Pr*(x_i *lowest*) to be estimated for all bidders, the reference bidder's cost estimate values value was substituted with the bid value in Friedman's parameter estimation procedure and in computing *p(i')*. Similarly, the constant x^*/c^* was modified to $x^*/x^* = 1$.

(3) Gates' method

Gates' approach is to estimate *Pr*(x_1 *lowest*) directly, by the formula:

$$\Pr(x_1\ lowest) = \left[\sum_{i=2}^{k} \frac{1-p(i')}{p(i')} + 1\right]^{-1} \qquad \textbf{\textit{(4)}}$$

where *p(i')* is the probability of the reference bidder's cost estimate being less than bidder *i*'s bid as described above. It can be shown that this is the exact result if, and only if, $f_i(x_i)$ is logistic, although a close approximation to the result may be obtained when normality and homogeneity (equal variances) are assumed.

Gates' model is recommended for use in situations where there is "sufficient bidding data relating to every competitor bidder on the particular job" (Gates, 1967:84). As with Friedman's method, a range of *q* values was tried and the same procedure used where less than *q* meetings occurred.

To enable *Pr*(x_i *lowest*) to be estimated for all bidders, cost estimates were again substituted by bid values.

(4) Carr's method

Carr uses Friedman's transformation for x_i and x_1 where, in addition, x_1 is also substituted by c_1, resulting in the unusual transformation

$$x_1^C = c_1/c_1 \qquad \textbf{(6)}$$

in which the denominator and numerator are assumed to be separate independent random variables. The arbitrary assumption that the x_i^F and x_1^C are normally and homogeneously (equal variances) distributed then allows the straightforward estimation of the required parameters from the frequency distribution of the pooled x_i^F ratios.

Carr's approach presents no difficulty in handling sparse data as normality and homogeneity are assumed. The only change made was to again substitute the cost estimates c_1 with the bids x_1 to enable all bidder's $Pr\ (x_i\ lowest)$ estimates to be made. Bidders with less than q data points were assigned equal probabilities (1/k).

(5) Skitmore's method

Skitmore's approach uses maximum likelihood estimates to fit the model $y_{ij} \sim N(\alpha_i+\beta_j, s_i^2)$ to the transformed values $y_{ij} = \mathbf{ln}(x_{ij}-mx_{(1)j})$ where m is a constant with 0.5<m<0.9 ($x_{(1)j}$ being the value of the lowest bid for the jth auction) and $N(\ )$ is the Normal probability density function. The probability prediction for y_1

$$\Pr(y_1) = \int_{-\infty}^{\infty} \frac{e^{-\frac{y_1^2}{2}}}{\sqrt{2\pi}} \cdot \prod_{i=2} \left[\int_{y_i = \frac{\sigma_1 y_1 + \mu_1 - \mu_i}{\sigma_i}}^{\infty} \frac{e^{-\frac{y_i^2}{2}}}{\sqrt{2\pi}} dy_i \right] dy_1 \qquad \textbf{\textit{(7)}}$$

is then obtained by substituting the estimates α_i and s_i into μ_i and σ_i respectively .

The method is difficult to apply in this form however as it involves the additional estimation of $x_{(1)}$. More recently, Skitmore and Pemberton's (1994) circumvented this problem by recourse to a simple natural log transformation, ie. $x_i^S = \boldsymbol{ln} x_i$, and this was also used here. From this, the probability of each bidder winning is computed. In cases where there was less than q data points for a bidder, that bidder was assigned an average α and s value.

The Skitmore model treats bidders with a single data point (one recorded previous bid) as having an alpha value based on that bid. A modification was also used which assigns these bidders with the average alpha value of all other bidders. The simple natural log transformation was again used.

DATA

Four sets of data, termed here Cases 1-4, were analysed. Cases 1-3 are reproduced in Appendix A.

Case 1 data

A construction company operating in the London area donated Case 1 data. The data covered all the company's building contract bidding activities during a twelve month period in the early 1980's in a total of 86 auctions. Details of the type of projects were available but not used in the analysis. Some of the data was incomplete, that is the value of some bids or the identity of bidders was not known by the company. In several cases, it was possible to supplement these data from a bidding information agency in the London area. The resulting number of auctions for which a full set of bids, together with the identity of the bidder, was available for analysis totalled 51.

Case 2 data

Case 2 data were donated by a north of England County Council for building contract bids over approximately four years prior to July 1982. Details of 258 contracts were provided in a precoded format. In other cases codes were missing or no tenders had been received. In yet other cases the codes or bids were illegible. The resulting number of contracts for which a full set of bids, together with the identity of the bidder, was available for analysis totalled 218.

Case 3 data

Case 3 data were obtained from the records of a bidding information agency in the London area. The agency held details of most bids for most building contracts in the London area in card form. A period of one week was spent copying a sample of contract data for the period

November 1976 to February 1977. The bids and associated bidders' names were recorded and the names later encoded for analysis. The resulting number of contracts for which a full set of bids, together with the identity of the bidders, were available for analysis totalled 373.

Case 4 data

Case 4 data comprised the combined Case 1 and Case 3 data. This was possible because of the identical bidder coding systems used for both Cases.

TESTING FRAMES

As the sole purpose of bidding models is for use in forecasting future outcomes, it is necessary to apply the models to out-of-sample data in addition to the in-sample data. Three forecast sample-testing frames are applied: (1) the *all-in* frame, (2) the *one-out* frame and (3) the *one-on* frame

The all-in frame

The *all-in* frame comprises all the data used to build the model. The testing procedure then simply tests the model against the data from which the model was built. In a similar way to the regression coefficient of determination, the error rate for *all-in* frame analysis is an unrealistically low measure of forecasting ability due to the self-fulfilling nature of the procedure.

The one-out frame

The *one-out* frame involves the use of the cross-validation procedure, by which the first auction is omitted from the model building process - the resulting model being applied to the omitted auction - this procedure being repeated, with replacement, for all the auctions in the Case set. Cross validation provides a reasonably realistic simulated out-of-sample test provided no significant time, or sequencing, effects are involved. The method is equivalent to the regression deleted residual analysis

The one-on frame

The *one-on* frame provides a simulation that is perhaps the closest to forecasting reality. A small sample of, say 13 auctions is used to build a model, which is then applied to the 14th auction - this procedure being repeated with 14 auctions to build the model which is applied to the 15th auction, etc. Of course, if the model performs better than chance, the results tend to improve as the number of auctions used to build the model increases. The final model, which incorporates all except the last auction, coincides with the final *one-out* model. The *one-out* frame results are therefore indicative of the final stages of the *one-on* results.

TESTS

Two fundamental approaches are available for testing the validity of a bidding model. One is to examine the structural correspondence of the model with its 'real-world' counterpart or prototype (Aris, 1978). The alternative, used here, is to test the model's predictions against actual outcomes. The first problem encountered in testing a model's probability prediction against an actual outcome is that the actual outcome of a single event in the bidding context is a binary result - one bidder wins (underbids all its competitors) or loses (is underbid by one of its competitors) the auction. For example, imagine an auction comprising several bidders and we are also testing two models, both of which provide a probability prediction of each bidder winning the auction. The highest ranked probability predictions for model 1 are bidder B, with 0.41, closely followed by bidder A, with 0.4. The highest ranked probability predictions for model 2 one the other hand are bidder A, with 0.3 and bidder, with 0.2. Assuming A actually wins the auction, which is the best model?

Intuitively, the best model is the one that provides the highest predicted probability for bidder A. This leads us to the conclusion that model 1 is best, as bidder A's predicted probability of winning (0.4) is greater than model 2’s predicted probability (0.3). However, it also seems intuitively reasonable that the model that best predicts the actual winner (i.e. the bidder with the highest relative probability of success) is best, i.e. model 2.

This suggests two distinct lines of analysis (1) to compare the *predicted binary win-lose outcome* with the actual binary win-lose outcome, and (2) to compare the *predicted probability* with the actual binary win-lose outcome.

Testing predicted binary outcomes: the binomial test

Here we wish to test the predicted binary win-lose outcome against the actual win-lose outcome. For example, consider a series of auctions comprising three bidders, A, B and C. Assume the model's probability predictions for auction 1 are 0.45, 0.35 and 0.2 for A, B and C respectively. We then convert these probability predictions to win-lose predictions, so that bidder A, having the highest probability prediction, is predicted to be the winner (and therefore, by implication, bidders B and C are predicted to be losers). The probability predictions for the other auctions are then processed in the same way. These win-lose predictions then compared with the actual winners and losers. Two questions now need to be answered (1) is the model doing any better than chance? and (2) given two models, which is the better model?

The binomial test can obviously be used to obtain the probability of a chance result for a set of k-size auctions by means of the cumulative binomial probability. Let R_k be a discrete binomially distributed random variable, then the probability of obtaining a value of r_k or less is the sum of the probabilities of obtaining a value of h=0, 1, ..., r_k, ie.,

$$\Pr(R_k \le r_k) = \sum_{h=0}^{r_k} \Pr(R_k = h) \qquad \textbf{\textit{(8)}}$$

the probability of obtaining each h for a k-size auction being

$$\Pr(R_k = h) = \binom{d_k}{h} k^{-h} \left(\frac{k-1}{k}\right)^{d_k - h} \qquad \textbf{\textit{(9)}}$$

over d_k auctions.

To combine different k-size auction sets, the required probability is

$$\Pr(R \leq r) = \sum_{h=0}^{r} \Pr(R = h)) \tag{10}$$

where

$$\Pr(R = h) = \sum_{m_2=0}^{d_2} \sum_{m_3=0}^{d_3} \ldots \sum_{m_t=0}^{d_t} \delta_h \prod_{w=1}^{t} \Pr(R_w = m_w) \tag{11}$$

d_2, d_3, ... , d_t denoting the number of k=2, 3, ... , t –size auctions respectively, and

$$\delta_h = 1 \textit{ when } m_2 + m_3 + \ldots + m_t = h, \textit{ otherwise } \delta_h = 0 \tag{12}$$

There are, however, logistical difficulties with ***(10)***. The computational time is excessive. A good approximation can be obtained by Case-wise simulation. Here, for s=1,2,…,s' iterations, *Pr(R=h)* in ***(11)*** is estimated by

$$\frac{1}{s'} \sum_{s=1}^{s'} \xi_h \tag{13}$$

where

$\xi_h = 1$ when $h_s = h$

$= 0$ otherwise

and

$$h_s = \sum_l \sum_i \gamma_{ijs} \qquad \textbf{\textit{(14)}}$$

where

$\gamma_{ijs} = 1$ when $P_{ij}^+ < 1/k_j$

$= 0$ otherwise

P_{ij}^+ being a random deviate from $U(0,1)$.

Testing predicted probabilities: the Logarithmic Score Test

Good's (1952) logarithmic score function, recently generalised by Bernado and Smith (1994:74-5), has been used previously in testing predicted probabilities of this kind (Wallace and Patrick, 1993; Dowe *et al*, 1996). This involves calculating the log scores, L, for each predicted probability, ie.,

$$L = 1 + \frac{\ln(p^*)}{-\ln\left(\frac{1}{2}\right)} \qquad \textbf{\textit{(15)}}$$

where $p^* = mod[P+(p-1)]$, P is the predicted probability and p is the outcome (p=1 if the bidder is the lowest, p=0 if the bidder is not the lowest).

To compare different models, the mean of L is calculated for each Case and *sample-frame*. High values indicate better predictions than low values. There are no statistical tests that indicate the significance of these differences.

RESULTS

Fig 1 shows the results for the binomial tests on the five models (**E**, **F**, **G**, **C**, **S1** and **S2** representing the Pim equal probability model, Friedman, Gates, Carr and the two Skitmore models respectively) for q=1,2, ..., 30 for Cases 2 and 4 for the *one-out* and *one-on frames*. This shows the quite different results that occur for each case and over the q value series. Table 3 summarises these results at the optimal q values for all the Cases and *frames*. The cumulative binomial probability, $Pr(R \leq r)$, is the probability the results could have occurred by chance alone and hence is around 0.5 for the **E** model. A $Pr(R \leq r)$ value of near zero therefore indicates a much superior performance than chance alone Thus, there are excellent predictive results for all models compared with **E** for the *all-in frame* for all Cases, and the *one-out* and *one-on frames* for Case 2. **C**, **S1** and **S2** models are also excellent for the *one-out frame* for Cases 3 and 4, with **C** and **S1** being very good for the *one-on frame* for Cases 3 and 4. Table 4 summarises the successful discrete (number of winners) predictions at the optimal q values of the various models. The first column contains the *k*-size of the contracts (*k*), second column gives the number of *k*-size contracts (*n*), **E1** is the real value of the random expectation, *n/k*, and **E2** is **E1** approximated to the nearest integer (0.5 being rounded upwards), **F**, **G**, **C**, **S1** and **S2** contain the number of wins successfully predicted by using the Friedman, Gates, Carr and Skitmore methods respectively, for Case 1. The other columns give the same statistics for Cases 2 to 4. For example, the all-in results for Case 1 show that there are six *k*=4-size auctions with an expected random (**E1**) winner prediction rate of 1.50 (6/*k*), which is two rounded to the nearest integer (**E2**). The Friedman (**F**) and Gates (**G**) models successfully predicted five out of the six winners, with the Carr (**C**) and original Skitmore (**S1**) models successfully predicting four winners and the modified Skitmore (**S2**) model successfully predicting three out of the six winners. The totals for all the *all-in* Case 1 results show the **F** and **G** models to be the best, with 37 winners being successfully predicted, followed by the **C**, **S1** and **S2** models, with 25, 23 and 16 successful predictions respectively. Ties sometimes occurred when a model predicted the same probability for two or more bidders in the same auction. The number of these tied results is given in Table 5.

Fig 2 shows the results for the logarithmic score tests for *q*=1,2, ..., 30 for Cases 2 and 4 for the *one-out* and *one-on frames*. As the figure shows, the models rarely better the equal probability model. Table 6 summarises these results at the optimal *q* values for all the Cases and *frames*, the best results being shown in bold text. The results indicate a general superiority of **S2** over the other models for the *one-out* and *one-on frames*. No statistical tests are available to test the significance of this result however.

DISCUSSION

In terms of straight lowest bidder prediction, the analysis shows the best models to be **C** and **S1**, with **S1** being marginally better than **C** for the larger sample sizes. The **F** and **G** models are generally much poorer. With the notable exception of the **F** and **G** model Case 3 and 4 *one-on frames,* all perform equal to or better than chance alone (model **E**). However, these results appear to be heavily influenced by the presence of ties – the weaker models producing a high number of tied predictions. It is clear that a mixed strategy, involving the random selection of predicted low bidders when ties are predicted, would have a considerable influence on these results.

The problem of ties does not of course exist with the Logscore analysis, as each bidder's probability prediction is taken in isolation. The results of this (Table 7) show the superiority of **S2** in six of the eight out of sample tests, being narrowly outperformed by **S1** in the Case 2 *one-on frame*.

One of the most striking aspects of the analysis is the close similarity of the results obtained by the **F** and **G** models in terms of predicted low bidders and wide dissimilarity in terms of actual probability estimates, the **F** model grossly underestimating the probabilities involved. This underestimation of probabilities is not unexpected and has been anticipated by several writers (e.g., Gates, 1967, Weverbergh, 1981, 1982) on theoretical grounds for it is plain that the expedient of assigning a zero variance to the reference bidder is certain to bias the results towards low probability estimates.

It is also unsurprising to find **S1** and **S2** producing the generally better results as these models, being multivariate, naturally utilise more of the available data.

CONCLUSIONS

Previous work in auction bidding has to a large extent been carried out without any real supporting data. In the context of construction contract auction bidding, it has been doubtful that sufficient data can be mustered for each bidder for any effective predictions to be made. In analysing some real and typical sets of auction bid data it has been possible to compare the major models here against pure chance and each other, showing that all offer an improvement on chance with multivariate models generally give the best results. The benefits if using these models in practice are, however, like the models themselves, statistical in nature. Like professional gamblers, proficient bidders given even a slight edge over chance should be able to exploit this to advantage over a period of time. The main drawback for construction contract bidders is that the precarious nature of the industry can mean that there is no long-term future as a single slip can often have terminal consequences.

Further empirical treatment of this topic would benefit by closer study of the behaviour of new entrants to aid the general development of predictive models where there is a shortage of data on individual bidders. The subject area would also benefit from the development of more sophisticated models that

- Enable the estimation of the covariance terms in the model to avoid the independence assumption

- Can handle a variety of uncertainties in the identity and/or number of bidders involved

- Allow the introduction of other influencing factors, such as contract type, size and location.

- Take account of changing market conditions and bidding behaviours over time

- Allow for different selection criteria or price award (e.g. the second lowest bid)

ACKNOWLEDGEMENTS

I am extremely grateful for the generous assistance of Professor Tony Pettitt in advising on the statistical tests needed for the data analysis, Dr Ingrid Baade for editing the first draft and to Dr H P Lo for his continual advice and encouragement..

REFERENCES

Aris, R., 1978, Mathematical modelling techniques, Pitman.

Bernado, J.M., Smith, A.F., 1994, Bayesian theory, John Wiley and Sons. ISBN 0-471 92416 4.

Carr, R.I., 1982, General bidding model, Journal of the Construction Division, Proceedings of the American Society of Civil Engineers, **108** (CO4), 639-50.

Dowe, L.D., Farr, Hurst, Lentin, 1996, Mathematics and computers in sport.

Friedman L 1956 A competitive bidding strategy, Operations Research **1**(4): 104-12.

Gates, M., 1967, Bidding strategies and probabilities, Journal of the Construction Division, Proceedings of the American Society of Civil Engineers, **93** (CO1), 75-107.

Good, I.J., 1952, Rational decisions, J. Roy. Statis. Soc. B, **14** 107-14.

McCaffer, R., 1976, Contractor's bidding behaviour and tender price prediction, PhD thesis, Loughborough University of Technology.

Park, W.R., 1972, Cost Engineering Analysis, J Wiley, New York.

Pim, J.C., 1974, Competitive tendering and bidding strategy, National Builder, **55** (11) pp. 541-545, **56** (2) pp. 56-57, **56** (10) pp.361-365, **57** (3) pp. 68-70.

Skitmore, R.M., 1991, The contract bidder homogeneity assumption: an empirical analysis, Construction Management and Economics, **9**(5) 403-29.

Skitmore, R.M., Pemberton, J., 1994, A multivariate approach to contract bidding mark-up strategies, J Opl Res Soc, **45**(11) 1263-72.

Wallace and Patrick, 1993, Machine Learning **11** 7-22.

Weverbergh, M., 1981, Competitive bidding models: an overview, Working paper 81-72, Aug, Centrum voor Bedrijfseconomie en Bedrijfseconometrie Universiteit Antwerpen - USFIA.

Weverbergh, M., 1982, Competitive bidding: estimating the joint distribution of bids, Working paper 82-79, Dec, Centrum voor Bedrijfseconomie en Bedrijfseconometrie Universiteit Antwerpen - USFIA.

Whittaker, J.D., 1970, A study of competitive bidding with particular reference to the construction industry, PhD thesis, City University, London.

SYMBOL TABLE

$\wedge(\ )$	lognormal pdf
δ_h	dummy variable
α_i	parameter in Skitmore's model
σ_{ij}	scale parameter for x_{ij}
β_j	parameter in Skitmore's model
μ_{ij}	location parameter for x_{ij}
γ_{ijs}	dummy variable
ξ_h	dummy variable
c	component of bid
c^*	value of bidder 1's cost estimate for the next auction
c_1	bidder 1's cost estimate
C	Carr's model
d_k	Binomial d for set of k-size auctions
E	Equal probability model
E1	real value of random expectation
E2	**E1** approximated to nearest integer
F	Friedman's model
$f(x_1,x_2,\ldots)$	joint pdf for bidders 1,2,…
$f_i(x_i)$	pdf of bidder i's bids
G	Gates' model
h	total number of correctly predicted lowest bidders for a set of auctions
h_s	comparator in Binomial test
i	bidder ID code
j	auction sequence number (j=1,2,…)
k	number of bidders in the auction
L	logarithmic score
m	constant in Skitmore's model
$m_2,m_3,\ldots$	components of h
$N(\ )$	normal pdf
p	outcome dummy variable in Logscore test
$p(i')$	probability bidder 1's cost estimate is lower than bidder i
p^*	logscore variable
P	predicted probability
P_{ij}^+	random deviate from $U(0,1)$
$Pr(\ )$	probability of being the lowest bidder
$Pr(x_1)$	probability bidder 1 will be the lowest
q	minimum data set for a bidder (**S1** and **S2**) or pair of bidders (**F** and **G**)
R	binomial R
r	binomial r
R_k	binomial R for set of k-size auctions
r_k	binomial r for set of k-size auctions
s	simulation number
s'	total number of simulations
s_i^2	parameter in Skitmore's model
S1	Skitmore's model
S2	modified Skitmore's model

t	total
$U(0,1)$	uniform pdf with min 0 and max 1
w	counter
$x_{(1)j}$	lowest bid in auction j
x^*	value of bidder 1's bid for the next auction
x_1^C	Carr's bid transformation (c_1/c_1)
x_i	bidder i's bid
x_i^F	Friedman's bid transformation (x_i/c_1)
X_i^F	random variable of which x_i^F is a value
x^S	Skitmore and Pemberton's bid transformation ($x^S = \mathbf{ln}x$)
$y_1, y_2, \ldots$	dummy variables in Equations ***(1)*** and ***(2)***
y_{ij}	Skitmore's bid transformation $\{y_{ij} = \mathbf{ln}(x_{ij} - mx_{(1)j})\}$

CAPTIONS

Table	**Caption**
1	*Frequency of low bids*
2	*Parameter estimation*
3	*Binomial probabilities*
4	*Predicted successes*
5	*Ties*
6	*Mean logscores*

Figure		**Caption**
1		Binomial tests *q* cut-offs
	a	Case 2 *one-out*
	b	Case 2 *one-on*
	c	Case 4 *one-out*
	d	Case 4 *one-on*
2		Logarithmic scores *q* cut-offs
	a	Case 2 *one-out*
	b	Case 2 *one-on*
	c	Case 4 *one-out*
	d	Case 4 *one-on*

TABLES

Company	No of auctions	No of bids	Mean no bids per auction	% win by chance	% actually won
A	41	249	6.1	16.4	17.0
B	36	183	7.0	14.2	15.4
C	19	88	4.6	21.6	21.1
D	35	202	5.8	21.6	17.1

Source: Pim (1974:541)

Table 1: Frequency of low bids

Source	**Parameters**	**Estimation**
Friedman	$F_i(x_i) \sim F_i(\mu_{i,}\sigma_i^{\ 2})$	x_i/c_1 ratios
	$F_1(x_1) \sim F_1(x^*/c^*{}_,0)$	
Gates	$F_i(x_i) \sim L(\ )$	Pr(x) estimated directly
	$F_1(x_1) \sim L(\ )$	
Carr	$F_i(x_i) \sim N(\mu_,\sigma^2/2)$	x_i/c_1 ratios
	$F_1(x_1) \sim N(1,\sigma^2/2)$	x_i/c_1 ratios
Skitmore	$F_i(x_i) \sim \wedge(\mu_{i,}\sigma_i^{\ 2})$	MLL (iteration)
	$F_1(x_1) \sim \wedge(\mu_1,\sigma_1^{\ 2})$	MLL (iteration)

N normal pdf
L logistic pdf
^ lognormal pdf

Table 2: Parameter estimation

Frame	Case 1		Case 2		Case 3		Case 4	
	Pr(R≤r)	*q*	*Pr(R≤r)*	*q*	*Pr(R≤r)*	*q*	*Pr(R≤r)*	*q*
All-in								
E	0.5103	-	0.4704	-	0.4286	-	0.5169	-
F	0.0000	0	0.0000	0	0.0000	0	0.0000	0
G	0.0000	0	0.0000	0	0.0000	0	0.0000	0
C	0.0000	0	0.0000	0	0.0000	0	0.0000	0
S1	0.0000	0	0.0000	0	0.0000	0	0.0000	0
S2	**0.0033**	**0**	**0.0000**	**0**	**0.0000**	**0**	**0.0000**	**0**
One-out								
E	0.5103	-	0.4704	-	0.4286	-	0.5169	-
F	0.2402	0	0.0006	3	0.4286	3	0.1771	7
G	0.1451	0	0.0002	3	0.4286	3	0.3332	7
C	0.2402	2	0.0002	30	0.0000	4	0.0000	0
S1	0.3651	0	0.0000	0	0.0000	0	0.0000	0
S2	0.3651	0	0.0000	0	0.0000	0	0.0000	0
One-on								
E	0.4863	-	0.4019	-	0.4608	-	0.5061	-
F	0.4863	5	0.0259	14	0.9973	6	0.9782	17
G	0.4863	5	0.0538	14	0.9973	6	0.9838	17
C	0.4863	0	0.0014	0	0.0302	0	0.0020	0
S1	0.4863	0	0.0002	0	0.0173	0	0.0008	0
S2	0.3239	0	0.0032	0	0.3621	0	0.0930	0

Table 3: Binomial probabilities

	Case 1								Case 2								Case 3								Case 4							
k	**n**	**E1**	**E2**	**F**	**G**	**C**	**S1**	**S2**	**n**	**E1**	**E2**	**F**	**G**	**C**	**S1**	**S2**	**n**	**E1**	**E2**	**F**	**G**	**C**	**S1**	**S2**	**n**	**R1**	**R2**	**F**	**G**	**C**	**S1**	**S2**
All-in results																																
2	0	0.00	-	-	-	-	-	-	23	11.50	12	19	19	15	19	19	40	20.00	20	34	34	23	28	27	40	20.00	20	34	34	22	29	27
3	1	0.33	0	0	0	1	1	0	16	5.33	5	15	15	11	13	12	53	17.67	18	47	47	31	35	32	54	18.00	18	47	47	32	31	32
4	6	1.50	2	5	5	4	4	3	27	6.75	7	17	19	16	13	12	47	11.75	12	38	38	19	24	18	54	13.50	14	43	43	23	25	25
5	5	1.00	1	3	3	3	3	1	31	6.20	6	21	22	10	14	11	59	11.80	12	53	54	26	28	26	64	12.80	13	55	55	26	28	26
6	21	3.50	3	15	15	7	8	6	47	7.83	8	34	34	19	21	17	85	14.17	14	74	74	42	36	35	105	17.50	17	87	87	47	47	45
7	9	1.29	1	6	6	4	3	4	35	5.00	5	19	17	10	10	8	48	6.86	7	44	44	18	18	14	57	8.14	8	50	50	21	22	19
8	5	0.63	1	5	5	4	2	1	20	2.50	3	14	14	4	8	5	28	3.50	4	25	26	12	9	7	33	4.13	4	27	28	13	12	8
9	3	0.33	0	2	2	1	1	1	12	1.33	1	6	5	3	3	3	8	0.89	1	7	7	4	5	4	11	1.22	1	10	10	4	5	4
10	1	0.10	0	1	1	1	1	0	2	0.20	0	2	2	2	1	1	4	0.40	0	4	4	4	2	2	5	0.50	0	5	5	5	3	2
11	0	0.00	-	-	-	-	-	-	2	0.18	0	2	1	0	0	0	1	0.09	0	1	1	0	0	1	1	0.09	0	1	1	0	0	0
12	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-
13	0	0.00	-	-	-	-	-	-	1	0.08	0	1	1	1	0	0	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-
14	0	0.00	-	-	-	-	-	-	2	0.14	0	2	2	1	1	1	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-
Total	*51*	*8.68*	*8*	*37*	*37*	*25*	*23*	*16*	*218*	*47.05*	*47*	*152*	*151*	*92*	*103*	*89*	*373*	*87.12*	*88*	*327*	*329*	*179*	*185*	*166*	*424*	*95.88*	*95*	*359*	*360*	*193*	*202*	*188*
One-out results																																
2	0	0.00	-	-	-	-	-	-	23	11.50	12	8	8	13	16	16	40	20.00	20	10	10	21	22	23	40	20.00	20	10	10	20	24	24
3	1	0.33	0	0	0	0	0	1	16	5.33	5	11	11	6	9	10	53	17.67	18	22	22	26	26	24	54	18.00	18	22	22	28	26	26
4	6	1.50	2	1	2	3	3	3	27	6.75	7	8	9	13	10	11	47	11.75	12	15	15	19	16	14	54	13.50	14	16	16	22	20	22
5	5	1.00	1	0	0	0	0	0	31	6.20	6	12	12	6	10	11	59	11.80	12	13	13	17	15	15	64	12.80	13	15	14	18	19	17
6	21	3.50	3	5	5	4	4	3	47	7.83	8	14	14	11	13	13	85	14.17	14	17	16	16	20	17	105	17.50	17	24	22	24	28	24
7	9	1.29	1	3	3	1	2	2	35	5.00	5	4	4	6	8	8	48	6.86	7	8	9	7	13	12	57	8.14	8	10	9	14	17	15
8	5	0.63	1	0	0	1	0	0	20	2.50	3	6	6	6	6	5	28	3.50	4	1	1	8	5	6	33	4.13	4	4	4	8	5	7
9	3	0.33	0	1	1	1	0	0	12	1.33	1	0	0	3	3	3	8	0.89	1	2	2	2	2	4	11	1.22	1	2	2	2	2	4
10	1	0.10	0	0	0	0	0	0	2	0.20	0	2	2	2	1	1	4	0.40	0	0	0	2	2	1	5	0.50	0	0	0	2	2	0
11	0	0.00	-	-	-	-	-	-	2	0.18	0	0	0	0	0	0	1	0.09	0	0	0	0	0	0	1	0.09	0	0	0	0	0	0
12	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-
13	0	0.00	-	-	-	-	-	-	1	0.08	0	0	0	0	0	0	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-
14	0	0.00	-	-	-	-	-	-	2	0.14	0	1	1	1	1	1	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-
Total	*51*	*8.68*	*8*	*10*	*11*	*10*	*9*	*9*	*218*	*47.05*	*47*	*66*	*67*	*67*	*77*	*99*	*373*	*87.12*	*88*	*88*	*88*	*118*	*121*	*116*	*424*	*95.88*	*95*	*103*	*99*	*138*	*143*	*139*
One-on results																																
2	0	0.00	-	-	-	-	-	-	23	11.50	12	6	6	15	16	13	38	19.00	19	4	4	15	20	17	38	19.00	19	4	4	20	22	17

3	1	0.33	0	0	0	0	0	0	14	4.67	5	9	9	7	7	6	49	16.33	16	8	8	18	21	17	50	16.67	17	9	9	18	21	17
4	3	0.75	1	2	2	2	2	3	27	6.75	7	8	8	9	6	7	46	11.50	12	10	10	15	15	16	53	13.25	13	13	13	19	20	20
5	5	1.00	1	0	0	0	0	0	30	6.00	6	7	7	7	9	9	59	11.80	12	10	10	14	11	11	64	12.80	13	12	11	17	14	14
6	15	2.50	2	2	2	1	3	1	47	7.83	8	13	13	14	13	11	83	13.83	14	14	14	17	17	14	103	17.17	17	20	20	23	23	19
7	7	1.00	1	1	1	2	1	2	33	4.71	5	7	7	5	8	9	46	6.57	7	9	9	6	7	5	55	7.86	8	11	11	9	10	8
8	5	0.63	1	0	0	0	0	0	17	2.13	2	1	1	3	2	2	27	3.37	3	3	3	4	4	5	32	4.00	4	3	3	6	6	7
9	2	0.22	0	1	1	1	0	1	11	1.22	1	2	1	2	2	2	8	0.89	1	4	4	3	2	1	11	1.22	1	4	4	3	2	1
10	1	0.10	0	0	0	0	0	0	1	0.10	0	1	0	1	1	1	4	0.40	0	0	0	1	1	0	5	0.50	0	0	0	1	1	0
11	0	0.00	-	-	-	-	-	-	1	0.09	0	0	0	0	0	0	1	0.09	0	0	0	0	0	0	1	0.09	0	0	0	0	0	0
12	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-
13	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-
14	0	0.00	-	-	-	-	-	-	2	0.14	0	2	2	0	1	1	0	0.00	-	-	-	-	-	-	0	0.00	-	-	-	-	-	-
Total	*39*	*6.53*	*6*	*6*	*6*	*6*	*6*	*7*	*206*	*45.15*	*46*	*56*	*54*	*63*	*65*	*61*	*361*	*83.79*	*84*	*62*	*62*	*98*	*100*	*86*	*412*	*92.55*	*92*	*76*	*75*	*116*	*119*	*103*

Table 4: Predicted successes

	Case 1					Case 2					Case 3					Case 4				
	F	**G**	**C**	**S1**	**S2**	**F**	**G**	**C**	**S1**	**S2**	**F**	**G**	**C**	**S1**	**S2**	**F**	**G**	**C**	**S1**	**S2**
All-in	2	2	0	0	3	5	4	1	0	1	19	18	1	1	11	17	17	1	1	8
One-out	4	4	0	0	3	15	15	12	1	3	69	69	2	1	10	59	59	2	2	10
One –on	2	2	1	1	7	23	23	6	6	14	118	119	16	13	48	118	120	16	13	48

Table 5: Ties

Frame	Case 1		Case 2		Case 3		Case 4	
	L	*q*	*L*	*q*	*L*	*q*	*L*	*q*
All-in								
E1	0.3516	-	0.2977	-	0.2638	-	0.2740	-
F	0.5919	0	0.5891	0	**0.7852**	0	0.7394	0
G	**0.6653**	0	**0.6528**	0	0.7800	0	**0.7668**	0
C	0.4249	0	0.5980	0	0.3422	0	0.3452	0
S1	0.4988	0	0.4321	0	0.3936	0	0.3944	0
S2	0.4280	1	0.4007	1	0.3669	0	0.3712	0
One-out								
E1	0.3516	-	0.2977	-	0.2638	-	0.2740	-
F	0.0965	7	0.1526	11	0.0425	5	0.0821	6
G	0.3499	19	**0.3223**	5	0.2051	4	0.2531	6
C	0.3497	29	0.3127	24	0.2688	25	0.2752	24
S1	0.3682	6	0.3084	2	0.2798	10	0.2884	12
S2	**0.3720**	5	0.3131	2	**0.2836**	9	**0.2908**	12
One-on								
E1	0.3578	-	0.2916	-	0.2657	-	0.2759	-
F	0.0738	8	0.0643	18	0.0778	3	0.0833	14
G	0.3506	19	0.2098	5	0.2473	3	0.2608	3
C	0.3747	30	0.2844	15	0.2690	29	0.2786	29
S1	0.3782	6	**0.2920**	30	0.2763	8	0.2853	8
S2	**0.3877**	5	0.2916	30	**0.2781**	8	**0.2864**	8

Table 6: Mean logscores